THE PHILOSOPHY AND GUIDELINES FOR REVISION OF THE THESAURUS OF ASTIA DESCRIPTORS

REPRINT OCTOBER 1964

1 NOVEMBER 1961

DEFENSE DOCUMENTATION CENTER
DEFENSE SUPPLY AGENCY

PHILOSOPHY OF AND GUIDELINES FOR REVISION OF THE ASTIA THESAURUS

1 November 1961

T. L. Gillum
ASTIA
P. H. Klingbiel
ASTIA
C. N. Mooers
Zator Company
E. Wall
Documentation Incorporated

Headquarters
Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Virginia

INTRODUCTION

In response to many inquiries from users of the Thesaurus of ASTIA Descriptors and from others interested in the use of controlled vocabularies for mechanized information retrieval, this paper is offered as an outline of the general plan to be followed in the preparation of a Second Edition of the ASTIA Thesaurus. The philosophies of the descriptor and thesaurus approaches to information retrieval are discussed, with particular emphasis on the relationships among descriptors. Although this document was intended as a guideline for individuals who had been invited to participate in the preparation of the Second Edition, the discussion of the thesaurus philosophy is believed to be of general interest to documentalists. Included is a bibliography which cites papers dealing with the general concept of technical vocabularies.

CONTENTS

BLANK PAGE

PHILOSOPHY OF AND GUIDELINES FOR REVISION OF THE ASTIA THESAURUS

I. INTRODUCTION

A. General Objectives

During the October 17-18, 1961, meeting of individuals and organizations interested in revision of the Thesaurus of ASTIA Descriptors, an ad hoc temporary committee on Thesaurus revision submitted a report (attached) containing a number of suggestions which appeared to meet with the approval of the assembled group. In accordance with these suggestions, this outline of the philosophy of, and guidelines for, Thesaurus revision is provided.

A major objective in revising the Thesaurus of ASTIA Descriptors is to provide an improved ASTIA indexing authority in a form most useful (1) to assist analysts in making consistent and sufficiently complete assignment of descriptors to accessioned technical information and (2) to assist bibliographers in making a corresponding consistent use of the descriptors during the formulation of inquiries for mechanized retrieval.

A second major objective in revising the Thesaurus is to create a device which will be as useful as possible to reference personnel in organizations other than ASTIA. In this connection, ASTIA is anxious during revision of the Thesaurus to have the cooperation and active participation of all individuals and organizations who can assist in making the Thesaurus more useful both to themselves and to ASTIA.

Edition II of the Thesaurus will incorporate the planned revisions. In addition, ASTIA is investigating means of notifying the users of its Thesaurus of subsequent changes, additions, and modifications to the Thesaurus. A number of alternative methods of performing this notification function are possible, and it is felt that this function will be easily performed -- particularly because the rate of Thesaurus modification in the future is expected to be low.

B. Definitions

To make meaningful the philosophy of ASTIA (as well as the herein contained guidelines for Thesaurus revision), certain terms and ideas must be defined.

1. Descriptors

Descriptors are controlled terms -- single words or phrases -- representing ideas or concepts. Descriptors are used to indicate the subject matter content of documents and technical information in other forms. Descriptors are to be distinguished from names of personal or corporate authors, from expressions giving contract numbers, and from other similar important kinds of access points of descriptive cataloging.

The word or phrase constituting each descriptor is chosen so that it will possess the maximum suggestiveness and convenience in indicating the descriptor's particular idea or concept to the technological or scientific group concerned. For example, the expression biological stains is more convenient to use than the inverted stains, biological in indicating the concept of this class of stains and of their use. Because several different functions are to be served by the descriptors, three broad types of descriptors are employed in the ASTIA Thesaurus.

a. Type A Descriptors

Type A descriptors are controlled or standarized names of subject-related sets of ideas or concepts. To describe them in another way, Type A descriptors represent very broad or generic concepts. In first approximation, they correspond to the names of the 292 descriptor groups included in Edition I of the Thesaurus of ASTIA Descriptors; e.g., acoustic detection. One intended purpose of the Type A descriptor of the revised Thesaurus is for broad classification of technical information in a compatible manner that will facilitate communication and exchange between information centers.

b. Type B Descriptors

Type B descriptors are the controlled and standardized names of suitably chosen single ideas or concepts. Type B descriptors are what are usually called merely "descriptors." They correspond to the approximately 7,000 descriptors in Edition I of the Thesaurus; e.g., sonar receivers. Some of these Type B descriptors may become Type A descriptors as a result of the current revision effort.

c. Type C Descriptors

Type C descriptors are terms extracted from the information being indexed to delineate information content not dealt with by Type A or B descriptors. Therefore, Type C descriptors are not completely controlled or standardized. The Type C descriptor terms must be specific in meaning. Ordinarily they will consist of proper or code names of equipment or projects, or will be important but infrequently used or parochial terminology; e.g.; AN/BQG-1. In ASTIA parlance, these terms are called identifiers (formerly known as "open-ended terms"). Type C descriptors provide additional and important points or access to ASTIA's document collection.

d. General Discussion

In analysis, descriptors of Type A and B are associated with each document (or other form of technical information) in order to delineate its subject matter. Type C descriptors are used as needed. Thus the document is delineated by a set of descriptors. To the extent

that it is practicable, each applicable and relevant descriptor of Type A and B in the Thesaurus is used to delineate any given document. In a fashion, the Thesaurus is used as a check list against the subject content.

An information search is prescribed by the formation of a small set of descriptors each of which is believed to be in the delineating set of the desired information. In the ideal case, selection occurs when a single small prescribing set is included in this delineating set. However, generally it will be necessary to use several prescribing sets to give the full range of selection needed.

One measure of the effectiveness of the revision of the ASTIA Thesaurus will be how closely it is possible to approach the ideal of a single prescribing set and a single inclusion for the search and selection of technical information, and for this set to be independent of the viewpoints of individual documentalists.

2. Relationships

The description of documents for effective retrieval is a communication process. An understanding of communications depends not only upon the terminology (i.e., descriptors) employed but also upon the context of that terminology as well as the meaning inferred by the recipient in the communication pattern. Context involves relationships among descriptor meanings -- and there exist several different kinds of relationships, which are discussed under Part II-A of this paper. If relationships among descriptors are not specified in a retrieval system, confusion as to descriptor meaning may develop during both input (analysis and indexing) and output (retrieval). On the other hand, the specification of relationships among descriptors enables consistent and sufficiently extensive use of the vocabulary.

3. Thesaurus

The ASTIA Thesaurus is an authoritative and structured reference to the ASTIA vocabulary of descriptors. As such, it exhibits the relationships among descriptors and their relationships to words in ordinary language, and clearly defines what sorts of relations exist among specific descriptors. This, in effect, assists greatly in defining each descriptor by relating it in specified fashions to other descriptors as well as to groups of descriptors and to common terminology.

As such, the Thesaurus of ASTIA Descriptors constitutes the basic "tool" by means of which ASTIA's objectives (of providing an authoritative vocabulary for consistent and extensive use by analysts and bibliographers) may be achieved with a high degree of simplicity and validity. This is the fundamental expression of ASTIA's philosophy.

C. Philosophy

The philosophy employed in constructing a vocabulary for ASTIA can be described from three viewpoints.

1. The Controlled Vocabulary

The vocabulary must be a controlled vocabulary. By "controlled" is meant that an authoritative and definitive reference is provided (i.e., the Thesaurus) both to descriptors and to relationships among descriptor meanings -- yet access to the vocabulary is possible from multiple viewpoints. The Thesaurus must also provide an authoritative guide from ordinary technical or scientific word usage to the controlled and standardized vocabulary of descriptors. Flexibility must be maintained (but disorder not permitted) with reference to the addition or deletion of descriptors as well as to addition or deletion of relationships among descriptors.

2. Competent Collection Coverage of the Vocabulary

The vocabulary must be competent to deal with the actual retrieval problems represented by the range, size, and depth of ASTIA's technical information collection. It must be useful in the processing of information and inquiries received by ASTIA. It must be expected to encompass only those technologies (or the terminologies thereof) encountered in the ASTIA collection. Yet, insofar as possible, it must be useful to other organizations dealing with collections dissimilar to ASTIA's.

3. Compatibility of the Vocabulary

Thus, the vocabulary should be as compatible as possible with other similarly-used vocabularies -- and the Thesaurus, as the principal means for achieving such compatibility, should make it possible for other organizations to "translate" their vocabulary to or from that of ASTIA -- and for ASTIA to do the same with other vocabularies. In this respect, the assistance of organizations other than ASTIA will prove invaluable.

II. GUIDELINES

A. Interdescriptor Relationships

The attainment of consistent and sufficiently extensive use of the ASTIA vocabulary during either input (analysis and indexing) or output (retrieval) operation depends upon overcoming three basic communications problems, listed here in increasing order of difficulty.

1. The Semantic Problem

This is the problem which may be narrowly defined as that of the meanings of words -- specifically, the relationship between the mental concept and the symbol which stands for that concept. In the following, a distinction will be made between words or phrases in ordinary language, which will be called terms, and the controlled and standardized expressions which we have been calling descriptors. In this narrow sense, there are three aspects to the semantic problem.

a. Homographs

Homographs are words which are spelled the same but which mean different things -- things not at all related, e.g., perch (bird roost) and perch (fish), tank (vehicle) and tank (container), lead (metal) and lead (electronic wiring component), etc. Such concepts must be distinguished one from the other or else consistency in document description and retrieval cannot be achieved.

b. Near-Synonyms

Depending upon viewpoint (see below) many terms may be synonymous or not. Some may even be synonyms from one viewpoint and antonyms from another; e.g., salvage (reclaiming) and recovery (reclaiming) vs. salvage (disposal) and recovery (reclaiming). The viewpoints used in defining descriptors for these concepts (e.g., salvage in the above example) must be made clear if consistency in document indexing and retrieval is to be achieved.

c. Synonyms

Cross references must be established for those terms which in ASTIA's environment are sufficiently near in meaning to descriptors such that item numbers are not sometimes posted to one descriptor and sometimes to another. However, care must be taken to insure that such definitions of synonymy are not made so broad that the fine detail of description is lost.

2. The Generic Problem

The generic problem involves the existence of "family trees" of concepts -- i.e., the broadness or narrowness of viewpoint brought to bear on a given concept. Terms standing for very narrow viewpoints of a concept tend to be Type C descriptors or identifiers (e.g., F4U, Minuteman, etc.). They will be very numerous but so specific that their utility is limited in a descriptor Thesaurus (as distinguished from their utility in retrieval).

However, there should exist another Thesaurus wherein these identifiers are referenced to the most specific or lowest generically related descriptors included in the descriptor Thesaurus -- e.g., F-106 (jet fighter). Identifiers must be cross-referenced among themselves to prevent confusion in and duplication of terminology; spelling must be standardized. Because identifiers are not under ASTIA internal control, full completeness and consistency cannot be expected at any stage.

Descriptors standing for broader viewpoints of a concept will be included in the descriptor Thesaurus. Each such descriptor will, when considered from any one viewpoint, be one member of a "generic tree". Consider, for example, the substance sodium chloride from the

chemical structure viewpoint. Salts include halides, sulfides, etc.; halides include chlorides, bromides, fluorides and iodides; chlorides include sodium chloride, aluminum chloride, etc. Here the term sodium chloride is a figurative leaf on the salts "generic tree" -- but this is true when it is considered from the chemical structure viewpoint. The same substance, when considered from the food viewpoint, would be a member of a "generic tree" containing the term seasoning agents. When considered from the refrigerant viewpoint, sodium chloride might even be generic to brine.

Each different viewpoint of the same concept will result in the concept being a member of a different generic family. Sodium chloride, for example, cannot always be considered as a seasoning agent, nor (for that matter) always as a refrigerant, an industrial raw material, a herbicide, etc. Rather, these are concepts which may be related to sodium chloride -- sometimes on the same generic level (i.e., nearly synonymous) and sometimes on different generic levels (i.e., members of the same generic family). Thus, in most instances, generic relationships cannot be specified among descriptors; variations in viewpoint make these relationships too transitory.

When a firm generic relationship exists among terms, that relationship must be exhibited in the Thesaurus; otherwise attempts at retrieval based upon either a broader or narrower consideration of the same viewpoint will fail. However, even though a firm generic relationship cannot be specified among certain descriptors, the possible existence of one must be exhibited in order to permit indexing and/or retrieval as necessary from various viewpoints.

3. Viewpoint Problem

This, the most difficult of the three basic problems, exhibits itself as facets of the semantic and generic problems of descriptors as described above. Thus the basic problems and their interrelationships can be diagrammed as follows:

	Degree of Variation of Viewpoint		
	Variations too frequent to permit specifying a relationship	Variations sufficiently infrequent, thus permitting specifying a defined relationship	Variations so marked (or so limited) as to make confusion unlikely
	(1)	(2)	(3)
Semantic Aspects	Near-synonyms or partial overlaps	Synonyms or almost complete overlap	Homographs (marked variation in viewpoint)
	(4)	(5)	(6)
Generic Aspects	Possible generic or inclusion relationships: (a) up (b) down	Defined generic or inclusion relationships: (a) up (b) down	Identifiers (limited variation in viewpoint)

B. Plan of Attack Upon Interdescriptor Relationship Problems

The above diagram thus defines six specific interdescriptor relationship problems, and the plan of attack upon each of these is set forth below.

1. Near Synonyms

Here there is a definite relationship between descriptors. The idea, concept, or meaning of one descriptor partially overlaps that of another descriptor, e.g., disposal, recovery, and salvage. Thus, for at least part of these meanings, we have different words for the same thing. The relationship in this example is only sometimes one of synonymy, but often is not, depending upon the variations in viewpoint. Thus the Thesaurus should indicate that there is a relationship or partial overlapping between certain of the descriptors, although the exact form of that relationship (or even its existence at all from some viewpoints) cannot always be specified.

The "Also See" reference is indicated in this circumstance. It must be from descriptor to descriptor. However, it must be recognized that the following exemplary condition may prevail:

Generators has "Also See" reference to motors.
Motors has "Also See" references to generators and to drives.
Drives has "Also See" reference to motors.

There may be no "Also See" reference between generators and drives because they may not be inherently related, although both may be related (from different viewpoints) to motors. Thus, one cannot expect all "Also See" references to be commutative, e.g., the "Also See" references of generators will not match exactly those of motors.

2. Synonyms

Here, variations among viewpoints (in the ASTIA environment) of two or more terms are adjudged to be so infrequent or so minor, and the difference in generic level is so minor, that a relationship of synonymy can easily be specified. Care must be taken not to specify synonymy when variations in viewpoint are so frequently encountered, or are so marked, as to make the specification untenable. The most frequently encountered "synonymous" descriptor should be used as the descriptor referred to from the "synonymous" terms used less frequently. The "Use" reference is indicated in this circumstance.

Terms affixed with "Use" references should definitely be inserted in their proper alphabetical order in the "Scope Note Index" (or its equivalent) of the Thesaurus -- as is done at present. If a term is a synonym (from two or more viewpoints) of more than one other descriptor, there is nothing wrong with a reference such as "Use Descriptor A or Descriptor B."

For purposes of future updating, it is advisable to provide "Includes" references, which would be affixed to descriptors "Used" in lieu of other terms; e.g., induction heating will have an "Includes" reference for every term which is referenced "Use induction heating."

3. Homographs

Treatment of these descriptors is simple, requiring only a "scope note" (such as the present parenthetical Descriptor Group name or other types of defining phrases).

4. Possible Generic Relationships

The same comments apply here as to the "near synonym" relationships (see above), except that variations in viewpoint affect the existence or absence of generic relationships rather than that of synonymy.

Here, too, use of the "Also See" reference is indicated, just as for the relationship of "near synonymy." The comment about noncommutativeness of the "Also See" references also applies here.

5. Defined Generic Relationships

Here, variations among viewpoints (in the ASTIA environment) of two or more descriptors may be adjudged to be so infrequent or so minor, while at the same time the difference in generic level is significant, that a generic relationship may be specified. Care must be taken not to specify a generic relationship when variations in viewpoint are so frequently encountered or are so marked as to make the specifications untenable.

ASTIA's philosophy of, and guidelines for, the treatment of generic relationships is as follows:

In terms of the known state-of-the-art, there does not exist a field-tested, automated system which solves the problems of indicating unambiguously vertical relationships for a multidiscipline library. This fact, true in May 1960 and true for November 1, 1961, explains why ASTIA exhibited no such relationships in the first edition of its Thesaurus. Because ASTIA's investment and expanding role in the better utilization of American scientific and technological know-how cannot be jeopardized, no presently known scheme, no matter how attractively or logically argued on paper, can be supported at this time.

Whatever the final disposition of the generic problem, any solution must be based on a controlled vocabulary. "Control" not only includes the authorization of a term as a descriptor and the definition of that term but also encompasses relationships among descriptors.

ASTIA proposes a system of indicating generic relationships which essentially treats of generic relationships among Type B descriptors.

In addition, ASTIA proposes to develop further techniques in generic indexing by prescribing relationships and usage between Type B and Type A descriptors.

Typically, a hierarchy will be created when the relation of generification is specified; e.g., all masers are microwave amplifiers, all microwave amplifiers are amplifiers, all amplifiers are electronic equipment, etc.

Use of the "Generic To" reference is indicated in this circumstance. For example, the descriptor microwave amplifiers would be referenced "Generic To masers" (as well as other types of microwave amplifiers covered by the vocabulary). The descriptor amplifiers would be referenced "Generic To microwave amplifiers, etc. (where "etc." refers to other types of amplifiers than microwave amplifiers as well as to specific kinds of both the other types of amplifiers and of microwave amplifiers). The descriptor electronic equipment would be referenced "Generic To amplifiers, microwave amplifiers, masers, etc." (where the "etc." includes other types of electronic equipment as well as all that was included by amplifiers).

In order to improve the thesaurus as a vocabulary reference tool, the standard dictionary practice of indicating the higher generic references is also recommended. For example, masers would be referenced "Add microwave amplifiers, amplifiers, electronic equipment." The descriptor microwave amplifiers referenced "Add amplifiers, electronic equipment." The descriptor amplifiers would be referenced "Add electronic equipment."

6. Identifiers

While these terms should not be part of the descriptor Thesaurus, each of them should be "tagged" with a descriptor, thus creating (in effect) an identifier Thesaurus. The "tags" should consist of higher generic levels of the concepts symbolized by the identifiers.

C. Procedures and Criteria for Selection and Deletion of Descriptors

Concepts to be expressed by Type B descriptors are selected (1) from accessioned technical information, (2) from bibliographic requests, and (3) by refinement of Type B descriptors which have been used frequently in processing information or requests.

In the first case, novel concepts which are thought to be candidates for Type B descriptors may be extracted from current documents and assigned as identifiers in order to determine their frequency of appearance (and corresponding utility as Type B descriptors) and to record the document numbers involved for updating of the retrieval tapes if the concept is subsequently incorporated into the Thesaurus.

In the second case, concepts which have not previously been recognized by assignment of Type B descriptors or identifiers may be revealed by users' questions and can be added when the pertinent documents are identified.

In the third case, Type B descriptors which are quite frequently used indicate (to some extent) concepts which may not be specific enough for efficient retrieval. Statistical studies of the assignment to documents of such descriptors are now made periodically to indicate which of them should be considered for refinement.

Suggestions for new descriptors (originating from all three of the aforementioned sources) are now evaluated in view of the logical, generic, and syntactical relationships to other descriptors; in view of definitions and the authority therefor; in view of the frequency with which the concept has appeared in the collection to date; and in view of the utility of the term in processing bibliographic requests. Decisions as to the actual descriptor terminology to be employed are based on the usage in textbooks, dictionaries, and other authoritative sources, as well as that found in the ASTIA collection. A descriptor proposal form (Attachment II) has been used with considerable success within ASTIA for evaluating descriptor suggestions.

Those descriptors which (1) experience has indicated to be too specific for efficient retrieval, (2) represent outmoded terminology, or (3) have been used very infrequently in processing current information and requests are candidates for deletion from the Thesaurus.

III. NON-THESAURUS CONSIDERATIONS

A. Names of Chemical Compounds

Although according to the previous discussion the names of chemical compounds might be treated as identifiers, it may happen that certain names of specific compounds should not be included even in the identifier Thesaurus. They should, of course, be "tagged" with the names of their "chemical families," which should be Thesaurus descriptors. Because a chemical compound will usually belong to more than one "chemical family," names of chemical compounds may thus turn out to be "exceptional" sorts of identifiers. This "exception" situation indicated that a different (possibly nonthesaurus) approach should be taken to the indexing of chemical compounds generally -- and this should be the object of a separate study.

B. Syntactical Problems

Only three basic problems (viewpoint, generic, and semantic) are discussed above. There is a fourth problem, that of syntax, which is relatively independent of the Thesaurus. Whether or not ASTIA should place syntactical constraints upon the descriptors is something that should be considered entirely apart from its studies of Thesaurus

revision at this time; any reasonable system of syntactical constraints will be compatible with any operationally successful Thesaurus.

Syntactical constraints would be employed principally to prevent "false drops" via preventing the invalid coordination of descriptors during retrieval. They can, however, also serve a useful purpose by making it possible to provide (in response to a search) not only a set of citation numbers ("addresses" of retrieved information) but also, for each citation, the descriptors associated with the information listed in ordered sequence.

Most frequently, role indicators are used as syntactical constraints, although (when average depth of indexing exceeds 30 to 40 descriptors per document) association links may be used as well. Role indicators provide clues as to the role a descriptor plays in the given document (e.g., raw material, production of, design of, research on, etc.); as such they enable the listing of descriptors in "ordered context." Association links are employed when the document being indexed is so complex that it must be indexed as if it were more than one document.

Role indicators must be few in number and (insofar as possible) mutually exclusive and collectively exhaustive. The design of a good set of role indicators is a major intellectual and experimental task not to be lightly undertaken. On the one hand, role indicators may be implicit anyway in some standard descriptor systems. On the other hand, the use of explicit role indicators may make the algebra of the entire process non-Boolean. However, once a role indicator system is designed, the use of the indicators will add only 10 to 20% to the cost of indexing and about the same amount to the size of the index (because about 10 to 20% of the descriptors assigned to each document will carry two roles). The use of association links, on the other hand, will add 50 to 150% to the cost of indexing and to the size of the index.

Finally, this kind of structural constraint, while specifically useful for certain aspects of chemical literature, may be quite inoperable in retrieval practice for many reports in other branches of science and technology.

IV. IMPLICATIONS OF OUTLINED PLANS AND GUIDELINES

The implementation of the aforementioned plans implies certain other actions which will result automatically. These are discussed below.

A. Multiword Descriptors

It is recognized that the inclusion in the Thesaurus of numerous multiword descriptors tends to increase the number of "Also See" references, to reduce the number of "Use" (and "Includes") references, and to have little effect upon the number of "Generic To" (and "Add") references. The net effect is to cause the Thesaurus to be physically larger, because

of the larger number of vocabulary descriptors generated by the various word combinations, but even more so because of the proliferation of "Also See" references (already the most numerous type of reference).

For this reason, during the revision of the Thesaurus, it is expected that some of the descriptors (represented by single or multiple words) will be "split" into descriptors of simpler (and broader) meaning. However, "splitting" must be avoided if the meaning of any part of the "split" descriptor is distorted from its meaning in the combined form, (e.g., half-life vs. half and life and air-to-surface vs. air and surface) or if the multiword descriptor is already heavily used. In this latter case, "splitting" of the heavily used multiword descriptor may result in the expenditure of excessive personnel and machine time to "recoordinate" these descriptors when servicing inquiries.

It should be noted that any steps to minimize the appearance of multiword descriptors in the Thesaurus need not prevent the operators of retrieval machinery from maintaining their own "precoordinations" of popular combinations of "split" terms for their own operational convenience.

B. Descriptor Group Redesign

It is recognized that many changes are desirable in the design of descriptor groups in order to implement the Type A descriptor concept. Some groups will be eliminated via absorption into existing groups; others will be eliminated by being split into newly defined groups; group names will be modified. ASTIA is already active in this work; however, it is expected that further modifications to descriptor groupings will result from an over-all examination of the results of the revision of the descriptor Thesaurus. This would, of course, be a task to be undertaken after the completion of the revision discussed in this paper.

V. IMPLEMENTATION OF PLANS

A. Independent Activities of ASTIA

As recognized by the aforementioned ad hoc committee, ASTIA is already proceeding with work leading to refinement of the Thesaurus. Aside from descriptor group redesign (see above) and routine maintenance work (addition of new descriptors, cross references, etc.), other non-routine activities are in progress and are described below.

1. Field Stabilization

This will involve the elimination of the present Field No. 13 (Miscellaneous Arts and Sciences) and the creation of a special pseudo-field to contain all general Type B descriptors which are not subject-matter equivalent to other Type B descriptors.

2. Low-Frequency Descriptor File

Low-frequency Type B descriptors are being evaluated for deletion of obsolete descriptors, and a file of about 2,000 low-frequency Type B descriptors is being created. This file will be used to perform manual searches. Bibliographic searches involving these descriptors are more quickly handled by hand than by machine.

3. High-Frequency Descriptor Refinement

ASTIA is investigating the descriptors with highest use frequency to determine whether these descriptors should be more precisely defined and whether the subject matter described by these descriptors can be better described by new, more specific descriptors.

4. General Plans

ASTIA plans to make available the appropriate amounts and quality of personnel and machine time to permit implementation of the Thesaurus revision program. In fact, people and machines are already active on the initial phases of this work.

B. Plans for Cooperative Activity

Two major activities are planned in which ASTIA invites participation by others. One of these is the development of an identifier Thesaurus -- a device to be created along the lines heretofore discussed. The other is the revision of the existing Thesaurus of ASTIA Descriptors.

1. Identifier Thesaurus

This work can being immediately. ASTIA invites active participation by small groups working in series or in parallel at ASTIA. The nature of this material permits groups with specialized interests to participate, because this material can be broken into such subject areas as chemistry, electronics, aeronautics, etc.

ASTIA believes that the construction of an adequate identifier Thesaurus is as important as the construction of a descriptor Thesaurus, and therefore rates this project as having 1-A priority.

2. Descriptor Thesaurus

The aforementioned ad hoc committee suggested that assistance by others to ASTIA in this particular effort should be provided by a small group of no more than about six people working at ASTIA with appropriate ASTIA personnel. The composition of this group need not be constant throughout the endeavor, but obviously the rotation into the task force of new members should be spaced to preserve a maximum continuity of experience.

The experience of ASTIA (and of others who have developed thesauri) has been that implementation of the "small task force" concept is both feasible and essential. Because of the great interest expressed by several groups representing the scientific and industrial communities, precedence will be given, in choosing the membership of the task force, to organizations which are engaged in thesaurus development. The task force will operate within the guidelines and principles outlined in this paper insofar as sound judgement and experience dictate.

VI. BIBLIOGRAPHY

1. Air Force Systems Command. February 1960. Vocabulary for Current ARDC Technical Efforts.

2. American Institute of Chemical Engineers. 1961. Chemical Engineering Thesaurus.

3. ASTIA. December 1959. Automation of ASTIA - A Preliminary Report. AD-227 000.

4. ASTIA. December 1960. Automation of ASTIA. AD-247 000.

5. ASTIA. October 1960. Controlling Literature by Automation. AD-243 000.

6. ASTIA. January 1961. Evolution of the ASTIA Automated Search and Retrieval System. AD-252 000.

7. ASTIA. February 1961. Guidelines for Using ASTIA Descriptors.

8. Bernier, C. L. 1957. Correlative Indexes II: Correlative trope indexes. Am. Document., 8(1):47-50.

9. Bernier, C. L. and K. F. Heumann. 1957. Correlative indexes III: Semantic relations among sematemes -- the technical thesaurus. Am. Document., 8(3):211-220.

10. Casey, R. S., J. W. Perry, A. Kent, and M. M. Berry. Punched Cards: Their Application to Science and Industry, 2nd ed., Reinhold Publishing Corporation, New York, 1958, pp 346-354.

11. Costello, J. C., Jr., and Eugene Wall. Recent Improvements in Techniques for Storing and Retrieving Information. In Coordinate Indexing, Volume V, Documentation, Inc.

12. Holm, B. E. and L. E. Rasmussen. 1961. Development of a technical thesaurus. Am. Document., 12(3):184-190.

13. Joyce, T. and R. M. Needham. 1958. The thesaurus approach to information retrieval. Am. Document., 9(3):192-197.

14. Klingbiel, P. H. Language Oriented Retrieval Systems. (To be announced as an ASTIA Document in December 1961.)

15. Masterman, M., R. M. Needham, and K. Sparck-Jones. November 1958. The analogy between mechanical translation and library retrieval. Presented at the International Conference on Scientific Information Storage and Retrieval.

16. Mooers, C. N. 1956. Zatocoding and developments in information retrieval. ASLIB Proceedings, 8(1):3-22.

17. Mooers, C. N. 1951. Zatocoding applied to mechanical organization of knowledge. Am. Document., 2(1):20-32.

18. Mooers, C. N. 1955. Information Retrieval on Structural Content. Third London Symposium, ed. by Colin Cherry. Royal Institution, London. Sept 12-16, 1955. p. 121. Butterworth Scientific Publications, London.

19. Newman, S. November 1958. Storage and retrieval of contents of technical literature, Patent Office Research and Development Reports No. 12.

20. Perry, J. W., Allen Kent and M. M. Berry. Machine Literature Searching. Interscience Publishers.

21. Taube, M., The preparation of manual dictionaries of association. In Coordinate Indexing, Vol. II, Documentation, Inc.

22. Wall, E. November 1959. A practical system for documenting building research. Presented at Building Research Institute.

REPORT OF THE COMMITTEE FOR DEVELOPMENT OF CRITERIA FOR GUIDELINES FOR THESAURUS REVISION

MEMBERS: C. N. Mooers, Chairman
E. B. Hincks
B. E. Holm
J. V. Philbrick
P. H. Klingbiel
A. J. Neumann

FOUR MAJOR POINTS ARE SUBMITTED:

I. Concurrence that ASTIA by 1 November 1961 prepare and circulate a first draft document covering the matters set forth in Item 6 of a list of recommendations* adopted by a thesaurus evaluation group which met 14-15 August 1961.

II. Recommendation that the following be considered as a suggested manner of preparing said document:

A. Deadline by 1. November 1961 for first draft.

B. The task force for the work should be members of ASTIA staff plus any outside personnel specifically assigned to the task or under contract.

C. ASTIA staff should be released for a preset number of hours per day, and given non-distracting quarters, for work on this draft.

D. ASTIA computer time should be made available for any quantitative assistance.

(Note: This Section II advisory only.)

III. Concurrence that bringing in working group members from outside ASTIA to work on the Thesaurus revision is to follow completion of said document. (This is not to preclude ASTIA from undertaking refining work immediately on the Thesaurus. It is merely that those on the outside want definite guidelines to follow before they can beneficially work on Thesaurus refinement.)

IV. Concurrence that the topics to be elaborated upon in this document include points A and B below:

*Item 6 of that meeting reads as follows: Preparation and publication of procedures, criteria, and standards for entry and deletion of retrieval terms; establish procedures for notification to users of Thesaurus changes on a periodic basis.

A. Develop criteria for different kinds of descriptors, their use, change, updating, deletion, etc. These three kinds of descriptors were discussed:

KIND A: Group Head Terms or their equivalent which may be suitable for interlibrary compatibility.

KIND B: Descriptors of the general kind now used in ASTIA.

KIND C: Terms similar to ASTIA Identifiers.

ASTIA is to draw up criteria on their distinction.

B. Delineate how to handle:

1. Relationships and cross references

2. Hierachies

3. Any other relationships

V. This Committee further concurred and advised as follows:

A. Descriptors of KIND B cannot be made completely compatible between libraries or from system to system. (Compatibility between systems will occur with KIND A primarily.)

B. The Thesaurus should be aimed specifically to be a tool for the librarian in a documentation center. (Descriptors of KIND A, however, should be usable by engineers for easy assignment to papers sent to other organizations.)

C. Assistance to ASTIA in the revision of the Thesaurus must be by a small outside group, such as 6 people or less. (Membership of the outside participating group may be rotating.)

D. The presumption must be made that machines will be more than capable to handle descriptor retrieval manipulations.

E. Terminology such as "Descriptor," "Keyword," etc. must be precisely defined in draft document and so used.

F. Support should be provided (machine time, programming) to utilize data already available such as descriptor frequencies, and the like.

CALVIN N. MOOERS
Chairman

18 October 1961

DESCRIPTOR PROPOSAL

1. Descriptor: Group No. Date:

2. Proposed cross references:
   Incl: Also See: Submitted by:
   Coordination:

3. Proposed definition:

4. Authority (literature citations and references to AD numbers):

5. Information on this subject presently contained in the AD collection might be retrieved by: